

Klasterisasi Tweet Terkait Dengan Pemilihan Presiden 2019 Menggunakan Ontology-based Concept Weighting dan DBSCAN

Puput Fajriati Tri Sholekah¹, Anisa Herdiani, S.T., M.T.², Ibnu Asror, S.T., M.T.³

^{1,2,3}Fakultas Informatika, Universitas Telkom, Bandung

¹puputfajriati@students.telkomuniversity.ac.id,

²anisaherdiani@telkomuniversity.ac.id, ³iasror@telkomuniversity.ac.id

1. Pendahuluan

1.1. Latar Belakang

Twitter merupakan salah satu media sosial yang sering digunakan oleh masyarakat Indonesia. Menurut Head of Business Development Twitter South East Asia and Australia, yang dinyatakan pada akun resmi twitter, Indonesia merupakan salah satu negara dengan jumlah pengguna Twitter terbesar di dunia [1]. Berdasarkan hasil survei tersebut, twitter merupakan suatu platform media sosial yang memiliki informasi beragam, salah satu contoh informasi beragam di twitter adalah *tweet* mengenai pemilihan presiden 2019 yang berhubungan dengan kedua calon pasangan presiden. *Tweet* yang beragam tersebut memiliki banyak topik pembahasan dan banyak yang tidak berkaitan dengan pemilihan presiden. Untuk menjadi sebuah topik pembahasan, *tweet* tersebut harus dikelompokkan agar dapat mengetahui topik apa saja yang berkaitan dengan calon pasangan presiden. Selain itu karena kebanyakan *user* twitter membuat *tweet* dengan bahasa yang tidak baku sehingga dapat menyebabkan data yang tidak beraturan terkait dengan pemilihan presiden 2019.

Berdasarkan masalah di atas, maka dibutuhkan sebuah sistem pengelompokan *tweet* untuk membedakan topik pembahasan mengenai pemilihan presiden yang terkait dengan kedua calon pasangan presiden. Sistem pengelompokan ini menggunakan metode dbscan yang dikombinasikan dengan *ontology-based concept weighting*. Topik pembahasan dalam penelitian ini ditandai dengan angka dan data *tweet* dikelompokkan berdasarkan kata yang sering muncul. Alasan menggunakan *ontology-based concept weighting* adalah untuk menerapkan pengetahuan tentang struktur hierarkis topik, jadi setiap topik yang berada pada tingkat hierarki yang sama, memiliki kesetaraan. Kelebihan dari *ontology* adalah membantu menemukan dokumen dengan kata-kata secara sintaksis berbeda tetapi memiliki makna yang sama [2]. Sedangkan, alasan menggunakan algoritma dbscan karena berdasarkan bentuk data yang tidak beraturan algoritma ini dapat menghasilkan *cluster* yang lebih akurat dibandingkan dengan *partitional clustering*. Kelebihan dari algoritma dbscan adalah dapat digunakan untuk mengelompokkan data dalam jumlah besar dan dapat menangani *noise/outlier* [3]. Setelah melakukan pengelompokan *tweet*, dilakukan perhitungan akurasi yang dinyatakan dalam bentuk *confusion matrix* dengan menggunakan evaluasi *F1 Score*.

1.2. Topik dan Batasan Masalah

Berdasarkan latar belakang, maka rumusan masalah dalam penelitian ini adalah mengelompokkan *tweet* terkait pemilihan presiden yang berhubungan dengan calon pasangan presiden, serta mengukur tingkat akurasi pengelompokan *tweet* menggunakan *ontology-based concept weighting* dan dbscan dengan menggunakan *F1 Score*.

Pada penelitian ini memiliki batasan masalah yang digunakan untuk membatasi penelitian agar berjalan sesuai dengan yang diharapkan. Batasan masalah pada penelitian ini adalah *tweet* yang digunakan adalah *tweet* yang berbahasa Indonesia dan tidak menamai *cluster* yang dihasilkan oleh algoritma dbscan.

1.3. Tujuan

Tujuan dari penelitian ini adalah untuk mengelompokkan *tweet* menggunakan *ontology-based concept weighting* dan dbscan yang dapat menjadikan ke dalam beberapa *cluster* sehingga menjadi sebuah informasi *tweet* tersebut termasuk ke dalam topik mana dan menunjukkan performansi pengelompokan *tweet* menggunakan *ontology-based concept weighting* dan dbscan.

1.4. Organisasi Tulisan

Laporan penelitian ini terdiri dari 4 bagian yaitu bagian 2 menjelaskan mengenai teori-teori yang berhubungan dengan penelitian yang dilakukan. Bagian 3 menjelaskan perancangan sistem Klasterisasi *Tweet* Terkait Dengan Pemilihan Presiden 2019 Menggunakan *Ontology-based Concept Weighting* dan DBSCAN. Pada bagian 4 dari penelitian ini adalah bagian evaluasi yang menjelaskan tentang hasil pengujian dan analisis hasil pengujian yang telah dilakukan. Sedangkan bagian 5 menjelaskan mengenai kesimpulan dan saran untuk mengembangkan sistem ini.